Immune cognition, social justice and asthma: structured stress and the developing immune system

Rodrick Wallace, PhD
Mindy Thompson Fullilove, MD
The New York State Psychiatric Institute
Deborah N Wallace, PhD
Division of Sociomedical Sciences
Joseph L Mailman School of Public Health
Columbia University*

August 13, 2001

Objective: We explore the implications of IR Cohen's work on immune cognition for understanding rising rates of asthma morbidity and mortality in the US. **Methods:** Immune cognition is conjoined with central nervous system cognition, and with the cognitive function of the embedding sociocultural networks by which individuals are acculturated and through which they work with others to meet challenges of threat and opportunity. Using a mathematical model, we find that externally-imposed patterns of 'structured stress' can, through their effect on a child's socioculture, become synergistic with the development of immune cognition, triggering the persistence of an atopic Th2 phenotype, a necessary precursor to asthma and other immune disease. **Conclusions:** Reversal of the rising tide of asthma and related chronic diseases in the US thus seems unlikely without a 21st Century version of the earlier Great Urban Reforms which ended the scourge of infectious

*Address correspondence to Rodrick Wallace, PISCS Inc., 549 W 123 St., Suite 16F, New York, NY, 10027. Telephone (212) 865-4766, email rdwall@ix.netcom.com.

diseases. **Key words:** American Apartheid, asthma, atopy, immune cognition, information theory, large deviation theory, renormalization, stress.

Introduction

Morbidity and mortality from asthma have risen nearly 50 % in the US since 1980 [1, 2]. In 1994 asthma-related demand for medical care accounted for one sixth of all emergency room visits and one out of eleven doctors' office visits [2]. By 1994, nearly 15 % of all urban children were afflicted with asthma, compared with 7 % of the entire US population. Among children one to four years of age, asthma hospital discharge rates increased 57 % between 1980 and 1992, nationally, and African-American children in this age range were six times more likely to die of asthma than Caucasian children [1].

Carr et al. [3] described the late 1980's geography of asthma in New York City: Minority neighborhoods such as Harlem, the South Bronx, Bedford-Stuyvesant, North Crown Heights and Washington Heights showed roughly five times the asthma mortality incidence of that found in affluent neighborhoods such as the Upper East Side, South Staten Island, and Forest Hills. Carr et al. [3] found that "Household income, percentage of population Black, and percentage of population Hispanic were significant predictors of area hospitalization rates (adjusted $R^2 = 0.75$)."

This pattern, which was later confirmed by DePalo et al. [4], is typical for other large US cities as well [5, 6], and suggests, *prima face*, that the highly structured psychosocial stressors of the system of American Apartheid have very recently become entrained into the developing immune systems of urban minority children.

We explore mechanisms by which such a 'phase transition' can take place, and, at the population level, produce widespread precursor conditions for a subsequent outbreaks of asthma and related atopic diseases.

Asthma is necessarily associated with failure of the child's developing immune system to switch from the Th2 'humoral' phenotype thought necessary to prevent maternal rejection *in utero* to a predominantly Th1 'cellular' phenotype more suited to the functioning of acquired immunity [7]. Biochemical feedback mechanisms tend to fix one or the other mode once it becomes developmentally predominant, although they tend to overlap somewhat, and are not 'orthogonal' [7]. This mechanism will be the focus of our modeling exercise.

According to current theory, five factors affect the 'decision' as to which phenotype will emerge from a newly-developed T cell, the 'naive' Th0 cell [7]. These include:

1. Local cytokine milieu; mainly IL-4 for Th2 and IL-12 for Th1.

2. Presence of immunologically active hormones, e.g., glucocorticoids stimulate Th2 and inhibit Th1. Similarly, catecholamines inhibit type 1 cytokine production and stimulate type 2 cytokines.

3. Dose and route of antigen presentation. High antigen doses suppress cell-mediated immunity, Th1, apparently a protective effect against self-tissue destruction.

4. The type of presenting cell stimulating the T cell.

5. The 'strength of signal' which is an ill-defined summation of the affinity of the T-cell receptor for the major histocompatibility complex (MHC), combined with the timing and density of receptor ligation.

Of these five, the cytokine milieu surrounding the newly-activated T cell is thought to be the most important, but this is profoundly influenced by the other four. That is, there appears to be a complicated 'grammar' and 'syntax' to a meaningful 'statement' which results in the binary outcome of the Th1/Th2 polarization. The nature of these 'statements' is profoundly affected by 'stress.'

As Elenkov and Chrousos [8] put the matter,

> "Recent evidence indicates that glucocorticoids and catecholamines, the end-products of the stress system... might selectively suppress cellular immunity [Th1], and favor humoral immune responses [Th2]. This is mediated by a differential effect of stress hormones and histamine on [Th1/Th2] patterns and type 1/type 2-cytokine production. Thus, systemically, stress might induce a Th2 shift, while, locally, under certain conditions, it might induce pro-inflammatory activities through neural activation of the peripheral corticotrophin-releasing factor-mast cell-histamine axis. Through the above mechanisms, stress may influence the onset and/or course of infectious, autoimmune/inflammatory, allergic and neoplastic diseases."

.

Clearly, then, the gestational and neonatal environment of the developing immune system will be critical in the 'decision' as to whether Th1 or Th2

immune phenotypes will predominate. Wright et al. [9] put the essential hypothesis as follows:

> "Prospective seroepidemiological studies have shown that the newborn period is dominated by Th2 reactivity in response to allergens, and it is also evident that the Th1 memory cells selectively develop shortly after birth (at 3-6 months of age) and persist into adulthood in non-atopic subjects. For most children who become allergic or asthmatic, the polarization of their immune systems into an atopic phenotype probably occurs during early childhood.
>
> These findings have sparked off vigorous investigation into the potential influence of early life environmental risk factors for asthma and allergy on the maturation of the immune system, in the hopes of understanding which factors will potentate (or protect from) this polarization... Although there is no direct evidence for the influence of stress on Th phenotype differentiation in the developing immune system, there is evidence that parental reports of life stress are associated with subsequent onset of wheezing in children between birth and one year. It has been speculated that stress triggers hormones in the early months of life which may influence Th2 cell predominance, perhaps through a direct influence of stress hormones on the production of cytokines that are thought to modulate the direction of differentiation."

The spatiotemporal pattern of the asthma increases among US children, in its exact match with patterns of residential segregation and community disintegration, suggests, however, that 'stress' is itself very highly structured. In this paper we will invoke Irun Cohen's theory of immune cognition to argue that the developing immune system interacts with, and is affected by, structured patterns of external stress through the intermediate medium of a local embedding – and cognitive – sociocultural network, of necessity including immediate family. Our development will further suggest that the internally coherent grammar and syntax, in a large sense, of that stress 'signal' are no less important than its 'magnitude'.

This is not an entirely new vision of the world. Recently, interactions between the central nervous system (CNS) and the immune system, and between the genetic heritage and the immune system have become officially

recognized and academically codified through journals with titles such as *Neuroimmunology* and *Immunogenetics*. Here we will argue that a cognitive socioculture – a social network embodying culture – in which individuals are embedded, and through which they are both acculturated and function to meet collective challenges of threat and opportunity, may interact strongly with individual immune function to produce a composite entity which might well be labeled an *Immunocultural Condensation* (ICC).

We first examine current visions of the interaction between genes and culture, and between the CNS and culture, and follow with a summary of Cohen's view of immune cognition. Next we argue that immune cognition and cognitive socioculture can become fused into a composite entity – the ICC – and that this composite, in turn, can be profoundly influenced by embedding systems of highly structured psychosocial and socioeconomic stressors. In particular, we argue that the internal structure of the stress – its 'grammar' and 'syntax' – are important in defining the coupling with the ICC.

The Appendix to this paper presents a detailed mathematical model of the ICC and its linkage with structured patterns of psychosocial or socioeconomic stress which is based on adapting renormalization techniques from statistical mechanics to information theory, in the spirit of the Large Deviations Program of applied probability. The necessity of such an approach will emerge from examination of IR Cohen's theory of immune cognition.

Genes, cognition, and culture

Increasingly, biologists are roundly excoriating simple genetic reductionism which neglects the role of environment. Lewontin [10], for example, explains that genomes are not 'blueprints,' a favorite public relations metaphor, as genes do not 'encode' for phenotypes. Organisms are instead outgrowths of fluid, conditional interactions between genes and their environments, as well as developmental 'noise.' Organisms, in turn, shape their environments, generating what Lewontin terms a triple helix of cause and effect. Such interpenetration of causal factors may be embodied by an array of organismal phenomena, including, as we shall discuss, culture's relationships with the brain and the immune system. We propose reinterpreting immune function in this light, in particular the coupling of the individual immune system with larger, embedding structures.

The current vision of human biology among evolutionary anthropologists is consistent with Lewontin's analysis and is summarized by Durham [11] as follows:

> "...[G]enes and culture constitute two distinct but interacting systems of inheritance within human populations... [and] information of both kinds has influence, actual or potential, over ... behaviors [which] creates a real and unambiguous symmetry between genes and phenotypes on the one hand, and culture and phenotypes on the other...
>
> [G]enes and culture are best represented as two parallel lines or 'tracks' of hereditary influence on phenotypes..."

With regard to such melding, over hominid evolution genes came to encode for increasing hypersociality, learning, and language skills, so the complex cultural structures which better aid in buffering the local environment became widespread in successful populations [12].

Every successful human population seems to have a core of tool usage, sophisticated language, oral tradition, mythology and music, focused on relatively small family/extended family groupings of various forms. More complex social structures are build on the periphery of this basic genetic/cultural object [13].

At the level of the individual human, the genetic-cultural object appears to be mediated by what evolutionary psychologists postulate are cognitive modules within the human mind [14]. Each module was shaped by natural selection in response to specific environmental and social conundrums Pleistocene hunter-gatherers faced. One set of such domain-specific cognitive adaptations addresses problems of social interchange [15]. The human species' very identity may rest, in part, on its unique evolved capacities for social mediation and cultural transmission. Anthropologist Robert Boyd has remarked that culture is as much a part of human biology as the enamel on our teeth.

Indeed, a brain-and-culture condensation has been adopted as a kind of new orthodoxy in recent studies of human cognition. For example Nisbett et al. [16] review an extensive literature on empirical studies of basic cognitive differences between individuals raised in what they call 'East Asian' and 'Western' cultural heritages. They view Western-based pattern cognition as 'analytic' and East-Asian as 'holistic.' Nisbett et al. [16] find that

1. Social organization directs attention to some aspects of the perceptual field at the expense of others.

2. What is attended to influences metaphysics.

3. Metaphysics guides tacit epistemology, that is, beliefs about the nature of the world and causality.

4. Epistemology dictates the development and application of some cognitive processes at the expense of others.

5. Social organization can directly affect the plausibility of metaphysical assumptions, such as whether causality should be regarded as residing in the field vs. in the object.

6. Social organization and social practices can directly influence the development and use of cognitive processes such as dialectical vs. logical ones.

Nisbett et al. [16] conclude that tools of thought embody a culture's intellectual history, that tools have theories build into them, and that users accept these theories, albeit unknowingly, when they use these tools.

We may assume, then, the existence of gene-culture and brain-culture condensations.

Immune cognition

Recently Atlan and IR Cohen [17] have proposed an information-theoretic adaptation of IR Cohen's [18, 19] 'cognitive principle' model of immune function and process, a paradigm incorporating pattern recognition behaviors analogous to those of the central nervous system.

Atlan and Cohen [17] describe immune system behaviors of cognitive pattern recognition-and-response as follows:

The meaning of an antigen can be reduced to the type of response the antigen generates. That is, the meaning of an antigen is functionally defined by the response of the immune system. The meaning of an antigen to the system is discernible in the type of immune response produced, not merely whether or not the antigen is perceived by the receptor repertoire. Because the meaning is defined by the type of response there is indeed a response repertoire and not only a receptor repertoire.

To account for immune interpretation IR Cohen [18] has proposed a cognitive paradigm for the immune system. The immune system can respond to a given antigen in various ways, it has 'options.' Thus the particular response we observe is the outcome of internal processes of weighing and integrating information about the antigen.

In contrast to Burnet's view of the immune response as a simple reflex, it is seen to exercise cognition by the interpolation of a level of information processing between the antigen stimulus and the immune response. A cognitive immune system organizes the information borne by the antigen stimulus

within a given context and creates a format suitable for internal processing; the antigen and its context are transcribed internally into the 'chemical language' of the immune system.

IR Cohen's cognitive paradigm suggests a language metaphor to describe immune communication by a string of chemical signals. This metaphor is apt because the human and immune languages can be seen to manifest several similarities such as syntax and abstraction. Syntax, for example, enhances both linguistic and immune meaning.

Although individual words and even letters can have their own meanings, an unconnected subject or an unconnected predicate will tend to mean less than does the sentence generated by their connection.

The immune system, in Atlan and Cohen's view, creates a 'language' by linking two ontogenetically different classes of molecules in a syntactical fashion. One class of molecules are the T and B cell receptors for antigens. These molecules are not inherited, but are somatically generated in each individual. The other class of molecules responsible for internal information processing is encoded in the individual's germline.

Meaning, the chosen type of immune response, is the outcome of the concrete connection between the antigen subject and the germline predicate signals.

The transcription of the antigens into processed peptides embedded in a context of germline ancillary signals constitutes the functional 'language' of the immune system. Despite the logic of clonal selection, the immune system does not respond to antigens as they are, but to abstractions of antigens-in-context.

Immune cognition and culture

As we show at length in the mathematical appendix, it is possible to give Atlan and Cohen's language metaphor of meaning-from-response a precise information-theoretic characterization, and to place that characterization within a context of recent developments which propose the 'coevolutionary' mutual entrainment – in a large sense – of different information sources to create larger metalanguages containing the original as subdialects [20-25]. This work, a highly natural extension of formalism based on the Large Deviations Program of applied probability, also permits treating gene-culture and brain-culture condensations using a similar, unified, conceptual framework of information source 'coevolutionary condensation'. Cohen's immune cognition

model suggests, then, the possibility that human culture and the human immune system may be jointly convoluted: That is, there would appear to be, in the sense of the gene-culture and brain-culture condensations of the previous section, an immune-culture condensation as well: To 'neuroimmunology' and 'immunogenetics' we add 'immunocultural condensation.'

The evolutionary anthropologists' vision of the world, as we have interpreted it, sees language, culture, gene pool, and individual CNS and immune cognition as intrinsically melded and synergistic. We propose, then, that culture, as embodied in a local cognitive sociocultural network, and individual immune cognition may become a joint entity whose observation may be 'confounded' – and even perhaps masked – by the distinct population genetics associated with linguistic and cultural isolation.

The 'decision' of the developing immune system to switch or not switch from a Th2 to a Th1 phenotype is significantly different from the minute-to-minute or day-to-day 'immediate' function mode of the immune system which Atlan and Cohen describe above: it takes place on a considerably longer time scale, over much of the first year of life. The 'chemical language' of immediate function must, then, be collapsed – 'integrated' – in some manner to form a sequence of chemical signals having a non-uniquely 'renormalized' grammar and syntax. That is, many different functional patterns of signal on a short time scale can give the same integral. The simplest hypothesis is that the integration or renormalization period, like so much else, is determined by the 24-hour human activity pattern, which suggests, for example, linkage of the child's developing immune system with the parental or familial cortisol-leptin cycle, which alternates over the day. Voice patterns, facial expression, pheromone emission, expressed emotion, and so on, may all play an immediate role.

The cortisol-leptin cycle is worthy of some comment: Leptin, the newly-discovered 'fat hormone', increases Th1 and suppresses Th2 cytokine production [26] and also stimulates proliferation and activation of circulating monocytes, and may play a direct role in inflammatory processes [27]. Leptin and cortisol have, however, a complex relation. Cortisol, an adrenal stress hormone, and leptin alternate their plasma peaks as part of the normal circadian cycle [28]. Cortisol increases can trigger answer leptin increases [29]. Glucocorticoid levels also influence plasma leptin levels [30]. Thus leptin and the adrenal hormones regulate each other: patterns of stress thus influence weight change, disease resistance, and inflammatory response. Th1/Th2 balance may be heavily influenced, in turn, by the adrenal hormone/leptin bal-

ance. Stress imposed on pregnant women may result in changes fetal immune and metabolic processes, with implications for birth weight, fat metabolism and risk for cardiovascular disease and allergenic susceptibility over the life course.

We suggest that the interplay of these factors over a day, and the correlational relations of renormalized or 'rate distorted' signals between sequences of days, constitutes no small part of the sociocultural milieu in which the child's developing immune system reaches its decision as to Th phenotype. The sociocultural network which envelops the child – including but not limited to parent or parents – in turn, engages in cognitive process to meet the structured challenges of threat and opportunity imposed upon it by the embedding socioeconomic system. Those challenges, to reiterate, have their own logical structure, their own grammar and syntax and, as the Appendix suggests, their powerful organization can impose itself down the nested hierarchy of interaction, to be translated, with some distortion, into the internal language of the child's developing immune system. As the next section indicates, this is not exactly a new thought.

Ecosystem theory: why neighborhoods count for individual health

Holling [31] has argued, since most ecologies are nested hierarchies, that a relatively few processes, having distinct frequencies in space and time, structure ecosystems, entrain other variables, and set the rhythm of ecosystem dynamics at other scales. A critical feature of such hierarchies is the asymmetric interactions between levels. In particular, the larger, slower levels maintain constraints within which the faster levels operate. In that sense, then, slower levels control faster ones, but, in the context of a loss of ecological resilience at the larger scale, faster processes can affect slower ones by means of their 'brittleness,' a concept explored at more length for human populations below and in several other papers of this series (e.g. [32]).

In Holling's view [31], ecosystems are structured hierarchically by a small number of underlying processes into a small number of levels, each characterized by a distinct scale of 'architectural' texture and of temporal speed of variables.

Each of the small number of processes influencing structure does so over limited scale ranges. The temporal and architectural structure of discrete ecosystem components are determined by three broad groups of processes, each dominant over different ranges of scale: *micro*, *meso* and *macro*.

According to Holling [31], the meso scale plays a particularly critical role. There, distinct disturbance phenomena are triggered at thresholds of tens of meters to kilometers. These driving variables of disturbance form the kind and amount of structure found at mesoscales by causing local events to cascade upward in scale to affect much larger landscape patterns. Analysis of the function of mesoscale process and structure thus provides the bridge between individual and global dynamics.

The mesoscale of human ecosystems is the neighborhood. Elsewhere we have explored how vulnerable neighborhoods of marginalized communities in the largest US cities constitute 'keystone populations,' in Holling's sense, for the national ecology of emerging and re-emerging infection (e.g. [33]), and more recently, in the spread of contagious behaviors associated with chronic disease [34]. Here we will extend this analysis downward in scale to examine some of the ways in which neighborhood structuring by the US system of de-facto urban Apartheid affects individual health and health behavior, in particular the decision of a developing immune system to adopt a predominant Th1 or Th2 phenotype.

Recent collapse of US minority urban neighborhoods

By 1980 large US urban minority neighborhoods – the keystones for public health at both larger and smaller scales – had begun to reflect a relentless siege by forces ranging from 'urban renewal', to contagious urban decay, 'planned shrinkage,' and the effects of deindustrialization. We recapitulate something of that history.

Urbanization of African Americans began at the start of the 20th Century. Small numbers migrated to Northern cities, and established themselves as "succession" communities in neighborhoods that had housed other ethnic groups seeking entry into American urban life. Gradually these urban communities expanded, incorporating later waves of immigrants, among them many who were forced out of agricultural work because of the mechanization of the farms. The integration of these newcomers into urban life was slow, yet the communities grew in complexity and organization, gaining political and economic power along the way [35].

As noted earlier, a series of policies hostile to poor urban neighborhoods undermined their physical and social infrastructures. In the most serious cases, urban renewal obliterated whole communities, which simply ceased to exist. Later policies of systematic disinvestment, including planned shrinkage, led to gradual destruction of individual buildings, and the outbreak of

widespread contagious urban decay which undermined wholesale the built environment. In either case, the functioning of social groups was fatally compromised by the alteration of the built environment within which they were embedded.

In the case of Harlem, New York, an important African American community, a stage/state model has been proposed to describe community disintegration [36]. The model assumes a range of possible states of community organization, from the highly integrated "model community" at one end, to the very disintegrated "collection of individuals" at the other. The stage/state model postulates that communities are not statically set at a single point on this range. Rather, they are dynamic entities, constantly working to maintain internal organization. External stressors, such as a war, loss of employment in the community, outbreaks of contagious urban decay or epidemic disease, can destabilize a community and trigger its decomposition. The Stage/State Model postulates that the transformation from "model" to "collection" follows a spiral pathway, in which each turn of the spiral is triggered by a destabilizing event.

Three turns of the spiral are proposed for Harlem. The first stage follows the initial loss of housing and is characterized by confusion in the population. Efforts were undertaken to reorganize the community, but the failure to rebuild the urban infrastructure created a barrier to full recovery. In the case of Harlem, further destabilization was caused by several factors, including the loss of manufacturing jobs and further loss of housing. This second stage was characterized by increasing disorder in the population. Along with the increase in disorder was an increase in the use and sale of psychoactive substances, and a decline in social controls on violence and related behaviors.

The growing use of drugs, licit and illicit, was accompanied by a shift in social relationships. Drug behaviors, which had been confined to designated areas, were able to occupy more and more territory. The effects on family life were magnified. The likelihood of family trauma, family dysfunction and family separation increased dramatically. Although in previous eras the 'home' and the 'street' had been carefully separated from each other, with increasing community disintegration, the 'street' was able to invade the territory of the 'home.'

This prepared the way for the incursion of crack cocaine in the mid-1980's [37-40]. The crack epidemic was extremely violent. Addiction to crack was very disabling, and involved a very large number of women. Both the noxious effects of crack, and the loss of the community-building efforts

of women, contributed to further destabilization of the community, a stage characterized by 'non-sense' in the population. In this stage, it was common to observe scenes that were frankly unreal by any measure of life prior to the arrival of crack cocaine.

At each level of scale – the personal, the family, the small group and the community – sharp changes in attitudes and behaviors have been documented. The exposure to trauma and violence has left psychological scars on a large portion of the population [41, 42]. The collapse of social relationships has increased the weight on individuals, as well as the sense of 'individualism.' The failure of group institutions has eroded the group power to contain and order behavior in public places. Further, behavior in public places has overwhelmed the interdiction to enter the home. The atomization of the population is one factor contributing to a marked loss of political participation and political consciousness among the population. The increasing political weakness of the group leaves it more vulnerable to majority decisions that are not in the best interest of the community. The siting of noxious facilities in vulnerable, disintegrating communities is one example.

The argument presented here suggests that the experience of history has altered the functioning of the socio-geographic community of Harlem, and created heavy social, political and emotional burdens for all residents. The future course of individual and group life will be build from the experience of disintegration: the downward spiral towards non-sense thus sets the pathways to the next stage.

The story of Harlem illustrates the experience of virtually every large African-American and Puerto-Rican urban community since the end of World War II. By 1980 this disruption had contributed materially to the milieu of 'structured stress' which has entrained immune development and function for many, if not most, community residents.

Discussion and conclusions

Our approach, implicitly, extends Cohen's vision of immune cognition one step by allowing day-to-day immune function to be 'renormalized' or 'integrated' into a longer-term developmental process, resulting in a 'higher order' cognitive decision regarding the binary shift from Th2 to Th1 immune phenotype. This is the 'initiating event' in the chain of causality resulting in asthma. It seems likely that a second iteration of our theory would be required to understand the 'promoting process' by which children with atopic

phenotype are driven to asthma. This work remains to be done, but we might speculate that, once a Th2 phenotype has been expressed, a second iteration could involve a leptin-cortisol-modulated or driven switch between phenotypes expressive of specific immunoglobin G (IgG) vs. specific immunoglobin E (IgE) antibodies, IgE being the most characteristic 'asthma antibody'.

In essence, we propose that an immunocultural condensation affects such 'higher order' processes as well, so that the cognitive functioning of the embedding sociocultural network becomes coevolutionarily condensed, in a large sense, with immune developmental cognition, and with the subsequent 'promotion' to explicit asthma. Although some mathematical details of such 'renormalized cognition' are discussed in the Appendix, further theoretical development will clearly be necessary.

A structured pattern of externally-imposed sociocultural or socioeconomic driving stressors has, at the population level, clear grammar and syntax: certain intercorrelated patterns of abuse and injustice are recognizable and 'make sense' within the system of American Apartheid, others do not. We have, for example, moved beyond mass killings and slave auctions, no small matter.

As the Appendix shows, in some detail, the population-level 'language' of externally-imposed stress can very suddenly become closely coupled with the 'language' of sociocultural cognition in a kind of 'phase transition' to form a coevolutionary condensation. It is this joint object which then forms the embedding milieu for a child's developing immune system.

We claim that policies and practices of urban renewal and planned shrinkage affecting urban minority communities had, by 1980, interacted with the outfalls of a massive deindustrialization driven in no small part by the diversion of engineering and scientific resources into the Cold War [e.g. 43]. This synergism greatly affected community sociocultural networks, triggering a literal phase transition in their function, so that the externally-imposed structured stress became quite literally a part of the 'higher order' immune cognition associated with development, and rapidly expressed itself in a rising proportion of ghetto children who develop the atopic phenotype. The mechanism may well enter into a feedback with similarly driven leptin-cortisol cycle dysfunctions leading to excessive weight gain.

Sociocultural phase transitions, like many such, are likely to be difficult to reverse, and may be subject to a path-dependent 'hysteresis', requiring far more than the simple removal of the triggering structured stress to reverse

[20-25]. Information-theoretic phase transitions may, then, tend to persist even when the conditions which triggered them are ameliorated. This is, essentially, because the internal grammar and syntax of an information source 'makes sense,' and the making-of-sense tends to be conserved in time and resist change. Elsewhere we have made a quantitative argument regarding such matters, in the character of 'generalized Onsager relations' [20-25]. Change, when it occurs, may, however, be very sudden. Thus corrective interventions against certain classes of problem – like the rise of asthma in minority urban communities – may not seem to work for a very long time, and will likely need to be highly proactive.

The foremost intervention, of course, would be to significantly change the system of 'structured stress' which is the driving force behind the 'asthma epidemic.' Programs, policies and practices which further destabilize urban minority communities will likely exacerbate asthma and related patterns of chronic disease.

The multiscale, hierarchical, and interactive nature of the urban ecosystem suggests that interventions against asthma will need to be similarly multiscale, hierarchical and interactive, with the most fruitful effort focused, not on the individual, but rather at the keystone neighborhood level.

We are suggesting, then, a 21st Century chronic disease version of the Great Urban Reform Movement of the late 19th and early 20th Centuries which brought a virtual end to the scourges of infectious disease – cholera, yellow fever, and the 'Capitan of all the Men of Death', tuberculosis.

Recent work suggests that, as with the earlier contagious scourges, the conditions leading to increases in chronic disease cannot be confined within marginalized urban communities. A large and growing body of work implies that Holling's vision of keystone communities is indeed correct, and that the social disintegration of urban minority neighborhoods is acting upward in scale as well as downward, entraining even suburban communities [44]. In particular the rising tide of asthma deaths has become regionalized in a very precise sense about certain US urban epicenters, as the growing decay indexed by the collapse of the inner-city itself became contagious, spreading into surrounding metropolitan areas along pathways defined by the daily journey-to-work [34]. Thus 'structured stress', once it emerged as a concentrated and proliferating social force within the largest central cities, could not be contained, and has already spread its deadly effect beyond the marginalized, dragging even the affluent down a slippery slope of chronic disease. Counterintuitively, the structural reforms needed to reverse the asthma epi-

demic within US central cities will benefit all, in much the same sense as did the bringing of clean water to the poor.

References

[1] Centers for Disease Control (CDC), 1996, Asthma mortality and hospitalization among children and young adults – United States, 1980-93. *Morbidity and Mortality Weekly Report*, **45**, 350-353.

[2] National Center for Health Statistics, 1996, Medical care for asthma increasing and changing, *Public Health Reports*, **87**, 112.

[3] Carr W, L Zeitel, K Weiss, 1992, Variations in asthma hospitalizations and deaths in New York City, *American Journal of Public Health*, **82**, 59-65.

[4] DePalo V, P Mayo, P Friedman and M Rosen, 1994, Demographic influences on asthma hospital admission rates in New York City, *Chest*, **106**, 447-451.

[5] Malveau F and S Fletcher-Vincent, 1994, Environmental risk factors for childhood asthma in urban centers, *Environmental Health Perspectives*, **103**, supplement 6, 59-62

[6] Liebman J, F Axler, L Kotranski and S Steinberg-Scribner, 1990, Letter to the editors, *New England Journal of Medicine*, **322**, 1606-1607.

[7] Spelberg B and J Edwards, 2001, Type 1/type 2 immunity in infectious diseases, *Clinical Infectous Diseases*, **32**, 76-102.

[8] Elenkov I and G Chrousos, 1999, Stress, cytokine patterns and susceptibility to disease, *Bailliere's Clinical Endocrinology and Metabolism*, **13**, 583-595.

[9] Wright R, M Rodriguez and S Cohen, 1998, Review of psychosocial stress and asthma: an integrated biopsychosocial approach, *Thorax*, **53**, 1066-1074.

[10] Lewontin R, 2000, *The Triple Helix: Gene, Organism and Environment*, Harvard University Press, Cambridge, MA.

[11] Durham W, 1991, *Coevolution: Genes, Culture and Human Diversity*, Stanford University Press, Palo Alto, CA.

[12] Bonner J, 1980, *The Evolution of Culture in Animals*, Princeton University Press, Princeton, NJ.

[13] Richerson P and R Boyd, 1995, The evolution of human hypersociality. Paper for Ringberg Castle Symposium on Ideology, Warfare and Indoctrinability (January, 1995), and for HBES meeting, 1995.

[14] Barkow J, L Cosmides and J Tooby (eds.), 1992, *The Adapted Mind: Biological approaches to mind and culture*, University of Toronto Press.

[15] Cosmides L and J Tooby, 1992, Cognitive adaptations for social exchange, in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, Oxford University Press, New York.

[16] Nisbett R, K Peng, C Incheol and A Norenzayan, 2001, Culture and systems of thought: holistic vs. analytic cognition, *Psychological Review*, **108**, 291-310.

[17] Atlan H and IR Cohen, 1998, Immune information, self-organization and meaning, *International Immunology*, **10**, 711-717.

[18] Cohen IR, 1992, The cognitive principle challenges clonal selection, *Immunology Today*, **13**, 441-444.

[19] Cohen IR, 2000, *Tending Adam's Garden: evolving the cognitive immune self*, Academic Press, New York.

[20] Wallace R and RG Wallace, 1998, Information theory, scaling laws and the thermodynamics of evolution, *Journal of Theoretical Biology*, **192**, 545-559.

[21] Wallace R and RG Wallace, 1999, Organisms, organizations and interactions: an information theory approach to biocultural evolution, *BioSystems*, **51**, 101-119.

[22] Wallace R, 2000, Language and coherent neural amplification in hierarchical systems: Renormalization and the dual information source of a generalized spatiotemporal stochastic resonance, *International Journal of Bifurcation and Chaos*, **10**, 493-502.

[23] Wallace R, 2000, Information resonance and pattern recognition in classical and quantum systems: toward a 'language model' of hierarchical neural structure and process,

`www.ma.utexas.edu/mp_arc-bin/mpa?yn=00-190`.

[24] Wallace R and R Fullilove, 1999, Why simple regression models work so well describing 'risk behaviors' in the USA, *Environment and Planning A*, **31**, 719-734.

[25] Wallace R, 2001, Selection pressure and organizational cognition: implications for the social determinants of health, `cogprints.soton.ac.uk`, Manuscript number cog00001526,

[26] Lord G, G Matarese, J Howard, R Baker, S Bloom and R Lechler, 1998, Leptin modulates the T-cell immune response and reverses starvation-induced immunosuppression, *Nature*, **394**, 897-901.

[27] Santos-Alvarez J, R Governa and V Sanchez-Margalet, 1999, Human leptin stimulates proliferation and activation of circulating monocytes, *Cellular Immunology*, **194**, 6-11.

[28] Bornstein S, J Licinio, R Tauchnitz, L Engelmann, A Negrao, P Gould and G Chrousos, 1998, Plasma leptin levels are increased in survivors of acute sepsis: associated loss of diurnal rhythm, in cortisol and leptin secretion, *Journal of Clinical Endicrinology and Metabolism*, **83**, 280-283.

[29] Newcomer J, G Selke, H Melson, J Gross, G Vogler and S Dagogo-Jack, 1988, Dose-dependent cortisol-induced increases in plasma leptin concentration in healthy humans, *Archives of General Psychiatry*, **55**, 995-1000.

[30] Eliman A, U Knutsson, M Gronnegard, A Steirna, K Albertsson-Wiklan and C Marcus, 1998, Variations in glucocorticoid levels within the physiological range affect plasma leptin levels, *European Journal of Endocrinology*, **139**, 615-620.

[31] Holling C, 1992, Cross-scale morphology, geometry, and dynamics of ecosystems, *Ecological Monographs*, **62**, 447-502.

[32] Wallace D and R Wallace, 2000, Life and death in Upper Manhattan and the Bronx: toward an evolutionary perspective on catastrophic social change, *Environment and Planning A*, **32**, 1245-1266.

[33] Wallace R and D Wallace, 1997, Community marginalization and the diffusion of disease and disorder in the United States, *British Medical Journal*, **314**, 1341-1345.

[34] D Wallace and R Wallace, 1998, Geography of asthma and diabetes over eight US metropolitan regions, *International Journal of Disease, Environment and Health Care*, **3**, 73-88.

[35] Fullilove MT, 1996, Psychiatric implications of displacement: contributions from the psychology of place, *American Journal of Psychiatry*, **153**, 1516-1523.

[36] Fullilove MT, 2000, AIDS in social context, in *AIDS-Related Cancers and their Treatment*, Feigal E, A Levine and R Abiggar, eds., Marcel Decker, New York, 371-386.

[37] Fullilove RE, M Fullilove, B Bowser, and S Gross, 1990, Risk of sexually transmitted disease among black adolescent crack users in Oakland and San Francisco, CA, *Journal of the American Medical Association*, **263**, 851-855.

[38] Fullilove MT, A Lown and R Fullilove, 1992, Crack 'hos and skeezers: traumatic experiences of women crack users, *Journal of Sex Research*, **29**, 275-287.

[39] Fullilove MT, R Fullilove, M Smith, K Winkler, C Michael, P Panzer, and R Wallace, 1993, Violence, trauma and post traumatic stress disorder among women drug users, *Journal of Traumatic Stress*, **6**, 105-115.

[40] Watkins B and MT Fullilove, 1999, Crack cocaine and Harlem, *Souls*, **1**, 37-48.

[41] Fullilove RE, MT Fullilove, M Northridge, M Ganz, M Bassett, D McLean, A Aidala, D Gemson, and C McCord, 1999, Risk factors for excess mortality in Harlem: findings from the Harlem Household Survey, *American Journal of Preventive Medicine*, **16**, 22-28.

[42] Fullilove MT, 1999, Root shock: the consequences of African American dispossession. Paper delivered at the New York Academy of Medicine Conference on Bioethics and the Health of the City, December 12.

[43] Wallace R, J Ullmann, D Wallace and H Andrews, 1999, Deindustrialization, inner-city decay and the hierarchical diffusion of AIDS in the US, *Environment and Planning A*, **31**, 113-139.

[44] Wallace D and R Wallace, 1998, *A Plague on Your Houses*, Verso Press, New York.

[45] Ash R, 1990, *Information Theory*, Dover Publications, New York.

[46] Cover T and J Thomas, 1991, *Elements of Information Theory*, Wiley, New York.

[47] Khinchine A, 1957, *The Mathematical Foundations of Information Theory*, Dover Publications, New York.

[48] Cohen IR, 2001, "Immunity, set points, reactive systems, and allograft rejection." To appear.

[49] Rojdestvenski I and M Cottam, 2000, "Mapping of statistical physics to information theory with applications to biological systems," *Journal of Theoretical Biology*, **202**, 43-54.

[50] Bennett C, 1988, "Logical depth and physical complexity". In: Herkin R (Ed) *The Universal Turing Machine: A Half-Century Survey*, Oxford University Press, pp. 227-257.

[51] Dembo A and O Zeitouni, 1998, *Large Deviations: Techniques and Applications, 2nd Ed.*, Springer-Verlag, New York.

[52] Granovetter M, 1973, "The strength of weak ties," *American Journal of Sociology*, **78**, 1360-1380.

[53] Wilson K, 1971, "Renormalization group and critical phenomena. I Renormalization group and the Kadanoff scaling picture", *Physical Review B*, **4**, 3174-3183.

[54] McCauley L, 1993, *Chaos, Dynamics and Fractals: an algorithmic approach to deterministic chaos*, Cambridge University Press, Cambridge, UK.

[55] Binney J, Dowrick N, Fisher A, Newman M, 1986, *The theory of critical phenomena*, Clarenden Press, Oxford.

[56] Atlan H, 1983, "Information theory," pp. 9-41 in R Trappl (ed.), *Cybernetics: Theory and Application*, Springer-Verlag, New York.

[57] Atlan H, 1987, "Self creation of meaning," *Psysica Scripta*, **36**, 563-576.

[58] Atlan H, 1998, "Intentional self-organization, emergence and reduction: towards a physical theory of intentionality," *Theses Eleven*, **52**, 6-34.

[59] Park H, S Amari and K Fukumizu, 2000, "Adaptive natural gradient learning algorithms for various stochastic models," *Neural Networks*, **13**, 755-764.

[60] Richerson P and R Boyd, 1995, "The evolution of human hypersociality." Paper for Ringberg Castle Symposium on Ideology, Warfare and Indoctrinability (January, 1995), and for HBES meeting, 1995.

[61] Richerson P and R Boyd, 1998, "Complex societies: the evolutionary origins of a crude superorganism," to appear.

[62] Luce R, 1997, "Several unresolved conceptual problems of mathematical psychology," *Journal of Mathematical Psychology*, **41**, 79-87.

Mathematical appendix

The essence of the modeling approach is to express the cognitive pattern recognition-and-response described as characterizing immune cognition by Atlan and Cohen [17] in terms of a 'language,' in a broad sense, and then to show how that language can interact and coalesce with similar cognitive languages at lager scales – central nervous system (CNS) and the embedding local sociocultural network. The next step is to demonstrate that these may, in turn, interact – and indeed coalesce –with a non-cognitive structured language of externally imposed constraint. The process of developing the formalism will make clear that change in such languages takes place 'on the surface,' in a sense, of what has gone before, so that path dependence – the persisting burden of history – becomes a natural outcome.

We begin with a summary of relevant information theory formalism.

Information theory preliminaries

Suppose we have an ordered set of random variables, X_k, at 'times' $k = 1, 2, ...$ – which we call $\mathbf{X}$ – that emits sequences taken from some fixed alphabet of possible outcomes. Thus an output sequence of length n, x_n, termed a path, will have the form

$$x_n = (\alpha_0, \alpha_1, ..., \alpha_{n-1})$$

where α_k is the value at step k of the stochastic variate X_k,

$$X_k = \alpha_k.$$

A particular sequence x_n will have the probability

$$P(X_0 = \alpha_0, X_1 = \alpha_1, ..., X_{n-1} = \alpha_{n-1}),$$

(1)

with associated conditional probabilities

$$P(X_n = \alpha_n | X_{n-1} = \alpha_{n-1}, ..., X_0 = \alpha_0).$$

(2)

Thus substrings of x_n are not, in general, stochastically independent. That is, there may be powerful serial correlations along the x_n. We call $\mathbf{X}$ an information source, and are particularly interested in sources for which the long run frequencies of strings converge stochastically to their time-independent probabilities, generalizing the law of large numbers. These we call *ergodic* [45-47]. If the probabilities of strings do not change in time, the source is called *memoryless*.

We shall be interested in sources which can be parametized and that are, with respect to that parameter, *piecewise memoryless*, i.e. probabilities do

not change markedly within a 'piece,' but may do so between pieces. This allows us to apply the simplest results from information theory, and to use renormalization methods to examine transitions between 'pieces.' Learning plateaus represent regions where, with respect to the parameter, the system is, to first approximation, memoryless in this sense. In what follows we use the term 'ergodic,' to mean 'piecewise memoryless ergodic.'

For any ergodic information source it is possible to divide all possible sequences of output, in the limit of large n, into two sets, S_1 and S_2, having, respectively, very high and very low probabilities of occurrence. Sequences in S_1 we call *meaningful*.

The content of information theory's Shannon-McMillan Theorem is twofold:

First, if there are $N(n)$ meaningful sequences of length n, where $N(n) \ll$ than the number of all possible sequences of length n, then, for each ergodic information source $\mathbf{X}$, there is a unique, path-independent number $H[\mathbf{X}]$ such that

$$\lim_{n\to\infty} \frac{\log[N(n)]}{n} = H[\mathbf{X}]. \tag{3}$$

See [2-4] for details.

Thus, for large n, the probability of *any* meaningful path of length $n \gg 1$ – independent of path – is approximately

$$P(x_n \in S_1) \propto \exp(-nH[\mathbf{X}]) \propto 1/N(n). \tag{3}$$

This is the *asymptotic equipartition property* and the Shannon-McMillan Theorem is often called the Asymptotic Equipartition Theorem (AEPT).

$H[\mathbf{X}]$ is the *splitting criterion* between the two sets S_1 and S_2, and the second part of the Shannon-McMillan Theorem involves its calculation. This requires introduction of some nomenclature.

Suppose we have stochastic variables X and Y which take the values x_j and y_k with probability distributions

$$P(X = x_j) = P_j$$

$$P(Y = y_k) = P_k$$

Let the joint and conditional probability distributions of X and Y be given, respectively, as

$$P(X = x_j, Y = y_k) = P_{j,k}$$

$$P(Y = y_k | X = x_j) = P(y_k | x_j)$$

The *Shannon uncertainties* of X and of Y are, respectively

$$H(X) = -\sum_j P_j \log(P_j)$$

$$H(Y) = -\sum_k P_k \log(P_j)$$

(4)

The *joint uncertainty* of X and Y is defined as

$$H(X, Y) = -\sum_{j,k} P_{j,k} \log(P_{j,k}).$$

(5)

The *conditional uncertainty* of Y given X is defined as

$$H(Y|X) = -\sum_{j,k} P_{j,k} \log[P(y_k|x_j)].$$

(6)

Note that by expanding $P(y_k|x_j)$ we obtain

$$H(X|Y) = H(X,Y) - H(Y).$$

The second part of the Shannon-McMillan Theorem states that the – path independent – splitting criterion, $H[\mathbf{X}]$, of the ergodic information source $\mathbf{X}$, which divides high from low probability paths, is given in terms of the sequence probabilities of equations (1) and (2) as

$$H[\mathbf{X}] = \lim_{n\to\infty} H(X_n|X_0, X_1, ..., X_{n-1}) =$$

$$\lim_{n\to\infty} \frac{H(X_0, ..., X_n)}{n+1}.$$

(7)

The AEPT is one of the most unexpected and profound results of 20th Century applied mathematics.

Ash [45] describes the uncertainty of an ergodic information source as follows;

> "...[W]e may regard a portion of text in a particular language as being produced by an information source. the probabilities $P[X_n = \alpha_n|X_0 = \alpha_0, ..., X_{n-1} = \alpha_{n-1})$ may be estimated from the

> available data about the language. A large uncertainty means, by the AEPT, a large number of 'meaningful' sequences. Thus given two languages with uncertainties H_1 and H_2 respectively, if $H_1 > H_2$, then in the absence of noise it is easier to communicate in the first language; more can be said in the same amount of time. On the other hand, it will be easier to reconstruct a scrambled portion of text in the second language, since fewer of the possible sequences of length n are meaningful."

Languages can affect each other, or, equivalently, systems can translate from one language to another, usually with error. The Rate Distortion Theorem, which is one generalization of the SMT, describes how this can take place. As IR Cohen [48] has put it, in the context of the cognitive immune system,

> "An immune response is like a key to a particular lock; each immune response amounts to a functional image of the stimulus that elicited the response. Just as a key encodes a functional image of its lock, an effective [immune] response encodes a functional image of its stimulus; the stimulus and the response fit each other. The immune system, for example, has to deploy different types of inflammation to heal a broken bone, repair an infarction, effect neuroprotection, cure hepatitis, or contain tuberculosis. Each aspect of the response is a functional representation of the challenge.
>
> Self-organization allows a system to adapt, to update itself in the image of the world it must respond to... The immune system, like the brain... aim[s] at representing a part of the world."

These considerations suggest that the degree of possible back-translation between the world and its image within a cognitive system represents the profound and systematic coupling between a biological system and its environment, a coupling which may particularly express the way in which the system has 'learned' the environment. We attempt a formal treatment, from which it will appear that both cognition and response to systematic patterns of selection pressure are – almost inevitably – highly punctuated by 'learning plateaus' in which the two processes can become inextricably intertwined.

Suppose we have a ergodic information source $\mathbf{Y}$, a generalized language having grammar and syntax, with a source uncertainty $H[\mathbf{Y}]$ that 'perturbs'

a system of interest. A chain of length n, a path of perturbations, has the form

$$y^n = y_1, ..., y_n.$$

Suppose that chain elicits a corresponding chain of responses from the system of interest, producing another path $b^n = (b_1, ..., b_n)$, which has some 'natural' translation into the language of the perturbations, although not, generally, in a one-to-one manner. The image is of a continuous analog audio signal which has been 'digitized' into a discrete set of voltage values. Thus, there may well be several different y^n corresponding to a given 'digitized' b^n. Consequently, in translating back from the b-language into the y-language, there will generally be information loss.

Suppose, however, that with each path b^n we specify an inverse code which identifies exactly one path $\hat{y}^n$. We assume further there is a measure of distortion which compares the real path y^n with the inferred inverse $\hat{y}^n$. Below we follow the nomenclature of [46].

The *Hamming distortion* is defined as

$$d(y, \hat{y}) = 1, y \neq \hat{y}$$

$$d(y, \hat{y}) = 0, y = \hat{y}.$$

(8)

For continuous variates the *Squared error distortion* is defined as

$$d(y, \hat{y}) = (y - \hat{y})^2.$$

(9)

Possibilities abound.

The distortion between paths y^n and $\hat{y}^n$ is defined as

$$d(y^n, \hat{y}^n) = (1/n) \sum_{j=1}^{n} d(y_j, \hat{y}_j)$$

(10)

We suppose that with each path y^n and b^n-path translation into the y-language, denoted $\hat{y}^n$, there are associated individual, joint and conditional probability distributions $p(y^n), p(\hat{y}^n), p(y^n, \hat{y}^n)$ and $p(y^n|\hat{y}^n)$.

The *average distortion* is defined as

$$D = \sum_{y^n} p(y^n) d(y^n, \hat{y}^n)$$

(11)

It is possible, using the distributions given above, to define the information transmitted from the incoming Y to the outgoing $\hat{Y}$ process in the usual manner, using the appropriate Shannon uncertainties:

$$I(Y, \hat{Y}) \equiv H(Y) - H(Y|\hat{Y}) = H(Y) + H(\hat{Y}) - H(Y, \hat{Y})$$

(12)

If there is no uncertainty in Y given $\hat{Y}$, then no information is lost. In general, this will not be true.

The *information rate distortion* function $R(D)$ for a source Y with a distortion measure $d(y, \hat{y})$ is defined as

$$R(D) = \min_{p(y|\hat{y}); \sum_{(y,\hat{y})} p(y)p(y|\hat{y})d(y,\hat{y}) \leq D} I(Y, \hat{Y})$$

(13)

where the minimization is over all conditional distributions $p(y|\hat{y})$ for which the joint distribution $p(y, \hat{y}) = p(y)p(y|\hat{y})$ satisfies the average distortion constraint.

The Rate Distortion Theorem states that $R(D)$, as we have defined it, is the maximum achievable rate of information transmission which does not exceed distortion D. Note that the result is *independent of the exact form of the distortion measure* $d(y, \hat{y})$.

More to the point, however, is the following: Pairs of sequences $(y^n, \hat{y}^n)$ can be defined as *distortion typical*, that is, for a given average distortion D, pairs of sequences can be divided into two sets, a high probability one containing a relatively small number of (matched) pairs with $d(y^n, \hat{y}^n) \leq D$, and a low probability one containing most pairs. As $n \to \infty$ the smaller set approaches unit probability, and we have for those pairs the condition

$$p(\hat{y}^n) \geq p(\hat{y}^n|y^n) \exp[-nI(Y, \hat{Y})].$$

(14)

Thus, roughly speaking, $I(Y, \hat{Y})$ embodies the splitting criterion between high and low probability pairs of paths. These pairs are, again, the input 'training' paths and corresponding output path.

Note that, in the absence of a distortion measure, this result remains true for two interacting information sources, the principal content of the *joint asymptotic equipartition theorem*, [46, Theorem 8.6.1].

Thus the imposition of a distortion measure results in a limitation in the number of possible jointly typical sequences to those satisfying the distortion criterion.

For the theory we will explore later – of pairwise interacting information sources – $I(Y, \hat{Y})$ (or $I(Y_1, Y_2)$ without the distortion restriction), can play the role of H in the critical development of the next section.

The RDT is a generalization of the Shannon-McMillan Theorem which examines the interaction of two information sources under the constraint of a fixed average distortion. For our development we will require one more iteration, studying the interaction of three 'languages' under particular conditions, and require a similar generalization of the SMT in terms of the splitting criterion for triplets as opposed to single or double stranded patterns. The tool for this is at the core of what is termed *network information theory* [46, Theorem 14.2.3]. Suppose we have (piecewise memoryless) ergodic information sources Y_1, Y_2 and Y_3. We assume Y_3 constitutes a critical embedding context for Y_1 and Y_2 so that, given three sequences of length n, the probability of a particular triplet of sequences is determined by *conditional probabilities with respect to* Y_3:

$$P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) =$$

$$\Pi_{i=1}^{n} p(y_{1i}|y_{3i}) p(y_{2i}|y_{3i}) p(y_{3i}).$$

(15)

That is, Y_1 and Y_2 are, in some measure, driven by their interaction with Y_3

Then, in analogy with the previous two cases, triplets of sequences can be divided by a splitting criterion into two sets, having high and low probabilities respectively. For large n the number of triplet sequences in the high probability set will be determined by the relation [46, p. 387]

$$N(n) \propto \exp[nI(Y_1; Y_2|Y_3)],$$

(16)

where splitting criterion is given by

$$I(Y_1; Y_2|Y_3) \equiv$$

$$H(Y_3) + H(Y_1|Y_3) + H(Y_2|Y_3) - H(Y_1, Y_2, Y_3)$$

Below we examine phase transitions in the splitting criteria H, which we will generalize to both $I(Y_1, Y_2)$ and $I(Y_1, Y_2|Y_3)$. The former will produce punctuated cognitive and non-cognitive learning plateaus, while the latter characterizes the interaction between selection pressure and sociocultural cognition.

Phase transition and coevolutionary condensation

The essential homology relating information theory to statistical mechanics and nonlinear dynamics is twofold [20-25]:

(1) A 'linguistic' equipartition of probable paths consistent with the Shannon-McMillan and Rate Distortion Theorems serves as the formal connection with nonlinear mechanics and fluctuation theory – a matter we will not fully explore here, and

(2) A correspondence between information source uncertainty and statistical mechanical free energy density, rather than entropy. See [20-25, 49] for a fuller discussion of the formal justification for this assumption, described by Bennett [50] as follows:

> "...[T]he value of a message is the amount of mathematical or other work plausibly done by the originator, which the receiver is saved from having to repeat."

This is a central insight: In sum, we will generally impose invariance under renormalization symmetry on the 'splitting criterion' between high and low probability states from the Large Deviations Program of applied probability [e.g. 51]. Free energy density (which can be reexpressed as an 'entropy' in microscopic systems) is the splitting criterion for statistical mechanics, and information source uncertainty is the criterion for 'language' systems. Imposition of renormalization on free energy density gives phase transition in a physical system. For information systems it gives interactive condensation.

This analogy is indeed a mathematical homology:

The definition of the free energy density for a parametized physical system is

$$F(K_1, ...K_m) = \lim_{V \to \infty} \frac{\log[Z(K_1, .., K_m)]}{V}$$

(17)

where the K_j are parameters, V is the system volume and Z is the 'partition function' defined from the energy function, the Hamiltonian, of the system.

For an ergodic information source the equivalent relation associates source uncertainty with the number of 'meaningful' sequences $N(n)$ of length n, in the limit

$$H[\mathbf{X}] = \lim_{n \to \infty} \frac{\log[N(n)]}{n}.$$

We will *parametize* the information source to obtain the crucial expression on which our version of information dynamics will be constructed;

$$H[K_1, ..., K_m, \mathbf{X}] = \lim_{n \to \infty} \frac{\log[N(K_1, ..., K_m)]}{n}.$$

(18)

The essential point is that while information systems do not have 'Hamiltonians' allowing definition of a 'partition function' and a free energy density, they may have a source uncertainty obeying a limiting relation like that of free energy density. Importing 'renormalization' symmetry gives phase transitions at critical points (or surfaces), and importing a Legendre transform in a 'natural' manner gives dynamic behavior far from criticality. Only the first of these will be addressed here. The second will likely be needed in any full theory of vaccine strategy.

As neural networks demonstrate so well, it is possible to build larger pattern recognition systems from assemblages of smaller ones. We abstract this process in terms of a generalized linked array of subcomponents which 'talk' to each other in two different ways. These we take to be 'strong' and 'weak' ties between subassemblies. 'Strong' ties are, following arguments from sociology [52], those which permit disjoint partition of the system into equivalence classes. Thus the strong ties are associated with some reflexive, symmetric, and transitive relation between components. 'Weak' ties do not permit such disjoint partition. In a physical system these might be viewed, respectively, as 'local' and 'mean field' coupling.

We fix the magnitude of strong ties, but vary the index of weak ties between components, which we call P, taking $K = 1/P$.

We assume the array, sensory activity and ongoing activity depend on three parameters, two explicit and one implicit. The explicit are K as above and an 'external field strength' analog J, which gives a 'direction' to the system. We may, in the limit, set $J = 0$.

The implicit parameter, which we call r, is an inherent generalized 'length' on which the phenomenon, including J and K, are defined. That is, we can write J and K as functions of averages of the parameter r, which may be quite complex, having nothing at all to do with conventional ideas of space, for example degree of niche partitioning in ecosystems.

Rather than specify complicated patterns of individual dependence or interaction for sensory activity, ongoing activity and array components, we follow the direction suggested above and instead work entirely within the domain of the uncertainty of the ergodic information source dual to the large-scale pattern recognition process, which we write as

$$H[K, J, \mathbf{X}]$$

Imposition of invariance of H under a renormalization transform in the implicit parameter r leads to expectation of both a critical point in K, which we call K_C, reflecting a phase transition to or from collective behavior across the entire array, and of power laws for system behavior near K_C. Addition of other parameters to the system, e.g. some Q, results in a 'critical line' or surface $K_C(Q)$.

Let $\kappa = (K_C - K)/K_C$ and take χ as the 'correlation length' defining the average domain in r-space for which the dual information source is primarily dominated by 'strong' ties. We begin by averaging across r-space in terms of 'clumps' of length R, defining J_R, K_R as J, K for $R = 1$. Then, following Wilson's [53] physical analog, we choose the renormalization relations as

$$H[K_R, J_R, \mathbf{X}] = R^{\mathcal{D}} H[K, J, \mathbf{X}]$$

$$\chi(K_R, J_R) = \frac{\chi(K, J)}{R}$$

(19)

where $\mathcal{D}$ is a non-negative real constant, possibly reflecting fractal network structure. The first of these equations states that 'processing capacity,' as indexed by the source uncertainty of the system which represents the 'richness' of the inherent language, grows as $R^{\mathcal{D}}$, while the second just states that the correlation length simply scales as R.

Other, very subtle, symmetry relations – not necessarily based on elementary physical analogs – may well be possible. For example McCauley, [54, p.168] describes the counterintuitive renormalization relations needed to understand phase transition in simple 'chaotic' systems.

For K near K_C, if $J \to 0$, a simple series expansion and some clever algebra [e.g. 20, 53, 55] gives

$$H = H_0 \kappa^{s\mathcal{D}}$$

$$\chi = \chi_0 \kappa^{-s}$$

(20)

where s is a positive constant. Some rearrangement produces, near K_C,

$$H \propto \frac{1}{\chi^{\mathcal{D}}}$$

(21)

This suggests that the 'richness' of the pattern recognition language is inversely related to the domain dominated by disjointly partitioning strong ties near criticality. As the nondisjunctive weak ties coupling declines, the efficiency of the coupled system as an information channel declines precipitously near the transition point: see [45-47] for discussion of the relation between channel capacity and information source uncertainty.

Far from the critical point matters are considerably more complicated, apparently driven by appropriate (and usually counterintuitive) generalizations of a physical system's 'Onsager relations'.

The essential insight is that *regardless of the particular renormalization symmetries involved, sudden critical point transition is possible in the opposite direction for this model*, that is, from a number of independent, isolated and fragmented pattern recognition systems operating individually and more or less at random, into a single large, interlocked, coherent pattern recognition system, once the parameter K, the inverse strength of weak ties, falls below threshold, or, conversely, once the strength of weak ties parameter $P = 1/K$ becomes large enough.

Thus, increasing weak ties between them can bind several different pattern recognition or other 'language' processes into a single, embedding hierarchical metalanguage which contains the different languages as linked subdialects.

This heuristic insight can be made exact using a rate distortion argument:

Suppose that two ergodic information sources $\mathbf{Y}$ and $\mathbf{B}$ begin to interact, to 'talk' to each other, i.e. to influence each other in some way so that it is possible, for example, to look at the output of $\mathbf{B}$ – strings b – and infer something about the behavior of $\mathbf{Y}$ from it – strings y. We suppose it possible to define a retranslation from the B-language into the Y-language through a deterministic code book, and call $\hat{\mathbf{Y}}$ the translated information source, as mirrored by $\mathbf{B}$.

Take some distortion measure d comparing paths y to paths $\hat{y}$, defining $d(y, \hat{y})$. We invoke the Rate Distortion Theorem's mutual information $I(Y, \hat{Y})$, which is a splitting criterion between high and low probability pairs of paths. Impose, now, a parametization by an inverse coupling strength K, and a renormalization symmetry representing the global structure of the system coupling. This may be much different from the renormalization behavior of the individual components. If $K < K_C$, where K_C is a critical point (or surface), the two information sources will be closely coupled enough to be characterized as condensed.

We will make much of this below; cultural and genetic heritages are generalized languages, as are neural, immune, and sociocultural pattern recognition.

Pattern recognition as language

The task of this section is to express cognitive pattern recognition-and-response in terms of a ergodic information source constrained by the AEPT. This general approach would then apply to the immune system, the CNS and sociocultural networks. Pattern recognition, as we will characterize it here, proceeds by convoluting an incoming 'sensory' signal with an internal 'ongoing activity' and, at some point, triggering an appropriate action based on a decision that the pattern of the sensory input requires a response. For the purposes of this work we do not need to model in any particular detail the manner in which the pattern recognition system is 'trained,' and thus adopt a 'weak' model which may have considerable generality, regardless of the system's particular learning paradigm, which can be more formally described using the RDT.

We will, fulfilling Atlan and Cohen's [17] criterion of meaning-from-response, define a language's contextual meaning entirely in terms of system output.

The model is as follows: A pattern of sensory input is convoluted with a pattern of internal 'ongoing activity' to create a path

$$x = (a_0, a_1, ..., a_n, ...).$$

This is fed into a (highly nonlinear) 'decision oscillator' which generates an output $h(x)$ that is an element of one of two (presumably) disjoint sets B_0 and B_1.

We take

$$B_0 = b_0, ..., b_k$$

$$B_1 = b_{k+1}, ..., b_m.$$

Thus we permit a graded response, supposing that if

$$h(x) \in B_0$$

the pattern is not recognized, and

$$h(x) \in B_1$$

that the pattern is recognized and some action $b_j, k+1 \leq j \leq m$ takes place.

We are interested in paths which trigger pattern recognition exactly once. That is, given a fixed initial state a_0 such that $h(a_0) \in B_0$, we examine all possible subsequent paths x beginning with a_0 and leading exactly once to the event $h(x) \in B_1$. Thus $h(a_0, a_1, ...a_j) \in B_0$ for all $j < m$ but $h(a_0, ..., a_m) \in B_1$.

For each positive integer n, let $N(n)$ be the number of paths of length n which begin with some particular a_0 having $h(a_0) \in B_0$, and lead to the condition $h(x) \in B_1$. We shall call such paths 'meaningful' and assume $N(n)$ to be considerably less than the number of all possible paths of length n – pattern recognition is comparatively rare – and in particular assume that the finite limit

$$H = \lim_{n \to \infty} \frac{\log[N(n)]}{n}$$

exists and is independent of the path x. We will – not surprisingly – call such a pattern recognition process ergodic.

We may thus define a ergodic information source **X** associated with stochastic variates X_j having joint and conditional probabilities $P(a_0, ...a_n)$ and

$P(a_n|a_0, ..., a_{n-1})$ such that appropriate joint and conditional Shannon uncertainties satisfy the relations

$$H[\mathbf{X}] = \lim_{n\to\infty} \frac{\log[N(n)]}{n}$$

$$= \lim_{n\to\infty} H(X_n|X_0, ..., X_{n-1})$$

$$= \lim_{n\to\infty} \frac{H(X_0, ...X_n)}{n+1}$$

We say this ergodic information source is *dual* to the pattern recognition process.

Different 'languages' will, of course, be defined by different divisions of the total universe of possible responses into different pairs of sets B_0 and B_1, or perhaps even by requiring more than one response in B_1 along a path. Like the use of different distortion measures in the RDT, however, it seems obvious that the underlying dynamics will all be qualitatively similar.

Meaningful paths – creating an inherent grammar and syntax – are defined entirely in terms of system response, as Atlan [56-58] and Atlan and Cohen [17] propose, quoting Atlan [57]

> "...[T]he perception of a pattern does not result from a two-step process with first perception of a pattern of signals and then processing by application of a rule of representation. Rather, a given pattern in the environment is perceived at the time when signals are received by a kind of resonance between a given structure of the environment – not necessarily obvious to the eyes of an observer – and an internal structure of the cognitive system. It is the latter which defines a possible functional meaning – for the system itself – of the environmental structure."

Elsewhere [20-25] we have termed this process an 'information resonance.'

Although we do not pursue the matter here, the 'space' of the a_j can be partitioned into disjoint equivalence classes according to whether states

can be connected by meaningful paths. This is analogous to a partition into domains of attraction for a nonlinear or chaotic system, and imposes a 'natural' algebraic structure which can, among other things, enable multitasking [20-25].

We can apply this formalism to the stochastic neuron: A series of inputs $y_i^j, i = 1...m$ from m nearby neurons at time j is convoluted with 'weights' $w_i^j, i = 1...m$, using an inner product

$$a_j = \mathbf{y}^j \cdot \mathbf{w}^j = \sum_{i=1}^{m} y_i^j w_i^j$$

(22)

in the context of a 'transfer function' $f(\mathbf{y}^j \cdot \mathbf{w}^j)$ such that the probability of the neuron firing and having a discrete output $z^j = 1$ is $P(z^j = 1) = f(\mathbf{y}^j \cdot \mathbf{w}^j)$. Thus the probability that the neuron does not fire at time j is $1 - f(\mathbf{y}^j \cdot \mathbf{w}^j)$.

In the terminology of this section the m values y_i^j constitute 'sensory activity' and the m weights w_i^j the 'ongoing activity' at time j, with $a_j = \mathbf{y}^j \cdot \mathbf{w}^j$ and $x = a_0, a_1, ...a_n, ...$

A little more work, described below, leads to a fairly standard neural network model in which the network is trained by appropriately varying the $\mathbf{w}$ through least squares or other error minimization feedback. This can be shown to, essentially, replicate rate distortion arguments, as we can use the error definition to define a distortion function $d(y, \hat{y})$ which measures the difference between the training pattern y and the network output $\hat{y}$ as a function of, for example, the inverse number of training cycles, K. As we will discuss in some detail, 'learning plateau' behavior follows as a phase transition on the parameter K in the mutual information $I(Y, \hat{Y})$.

Park et al. [59] treat the stochastic neural network in terms of a space of related probability density functions $[p(\mathbf{x}, \mathbf{y}; \mathbf{w})|\mathbf{w} \in \mathcal{R}^m]$, where $\mathbf{x}$ is the input, $\mathbf{y}$ the output and $\mathbf{w}$ the parameter vector. The goal of learning is to find an optimum $\mathbf{w}^*$ which maximizes the log likelihood function. They define a loss function of learning as

$$L(\mathbf{x}, \mathbf{y}; \mathbf{w}) \equiv -\log p(\mathbf{x}, \mathbf{y}; \mathbf{w}),$$

and one can take as a learning paradigm the gradient relation

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \partial L(\mathbf{x}, \mathbf{y}; \mathbf{w})/\partial \mathbf{w},$$

where η_t is a learning rate.

Park et al. [59] attack this optimization problem by recognizing that the space of $p(\mathbf{x}, \mathbf{y}; \mathbf{w})$ is Riemannian with a metric given by the Fisher information matrix

$$G(\mathbf{w}) = \int \int \partial \log p/\partial \mathbf{w} [\partial \log p/\partial \mathbf{w}]^T p(\mathbf{x}, \mathbf{y}; \mathbf{w}) d\mathbf{y} d\mathbf{x}$$

where T is the transpose operation. A Fisher-efficient on-line estimator is then obtained by using the 'natural' gradient algorithm

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t G^{-1} \partial L(\mathbf{x}, \mathbf{y}; \mathbf{w})/\partial \mathbf{w}.$$

Again, through the synergistic family of probability distributions $p(\mathbf{x}, \mathbf{y}; \mathbf{w})$, this can be viewed as a special case – a 'representation', to use physics jargon – of the general 'convolution argument' given above.

Again, it seems that a rate distortion argument between training language and network response language will nonetheless produce learning plateaus, even in this rather elegant special case.

The foundation of our mathematical modeling exercise is to assume that both the immune system and a sociocultural network's pattern recognition behavior, like that of other pattern recognition systems, can also be represented by the language arguments given above, and is thus dual to a ergodic information source, a context-defining language in Atlan and Cohen's sense [17], having a grammar and syntax such that meaning is explicitly defined in terms of system response.

Sociogeographic or sociocultural networks – social networks embedded place and embodying culture – serve a number of functions, including acting as the local tools for teaching cultural norms and processes to individuals. Thus, for the purposes of this work, a person's social network – family and friends, workgroup, church, etc. – becomes the immediate agency of cultural dynamics, and provides the foundation for the brain/culture 'condensation' whose pathologies are analyzed by R Wallace and M Fullilove [25].

Sociocultural networks serve also, however, as instruments for collective decision-making, a cognitive phenomenon. Such networks serve as hosts to a political, in the large sense, process by which a community recognizes and responds to patterns of threat and opportunity. To treat pattern recognition on sociocultural networks we impose a version of the structure and general formalism relating pattern recognition to a dual information source:

We envision problem recognition by a local sociocultural network as follows: A 'real problem,' in some sense, becomes convoluted with a community's internal sociocultural 'ongoing activity' to create the path of a 'perceived problem' at times $0, 1, ...$, producing a path of the usual form $x = a_0, a_1, ..., a_n, ...$ That serially correlated path is then subject to a decision process across the sociocultural network, designated $h(x)$ which produces output in two sets B_0 and B_1, as before. The problem is officially recognized and resources committed to if and only if $h(x) \in B_1$, a rare event made even more rare if resources must then be diverted from previously recognized problems.

For the purposes of this work, then, we will view 'culture' as, in fact, a sociocultural cognitive process which can entrain individual cognition, a matter on which there is considerable research [e.g. 60, 61].

Our next task is to apply phase transition dynamics to ergodic information sources dual to a pattern recognition language, using techniques of the sections above. Similar considerations will apply to 'non-cognitive' interaction between structured selection pressures and the affected system.

To reiterate, we have, following the earlier discussion of Atlan and Cohen's work [17], implicitly assumed that the immune cognition can likewise be expressed as a pattern recognition-and-response language characterized by an information source uncertainty.

'Higher order' pattern recognition

Suppose the pattern recognition language of the previous section is, in some sense, iterated, so that its output sequence of the decision oscillator $h(x) \in B_0, B_1$ serves as the input to a higher order decision oscillator, h^b having, for example, a simple binary output: $\mathrm{h}^b = Th1$ or $h^b = Th2$. Thus the sequence of $x^b \equiv b_0, b_1, ..., b_k, ...$ becomes the object of interest, leading to the binary output. In all likelihood, however, the output of h will be condensed by integration or renormalization, as discussed before. Thus the x^b will first be mapped or coded onto a simplified alphabet. In any event,

the *original* paths in $x = a_0, a_1, ..., a_k, ...$ can be placed in equivalence classes, according to those which, ultimately, after iteration, lead to outputs Th2 or to Th1.

The process can, of course, be iterated to a higher level, so that, if the switdh is indeed thrown, the Th2 phenotype paths will become input for another decision oscillator, leading to outputs of no disease, asthma, or some other atopic dysfunction.

Obviously, at this point mathematical possibilities begin to grow exponentially, including the necessity of examining 'meaningful' binary sequences of internal and external signals, and it seems likely that recourse to empirical study will be required to prune the thicket.

Generalized cognitive condensations

We suppose a cognitive system – more generally a linked, hierarchically structured, and broadly coevolutionary condensation of several such systems – is exposed to a structured pattern of sensory activity – the training pattern – to which it must learn an appropriate matching response. From that response we can infer, in a direct manner, something of the form of the excitory sensory activity. We suppose the training pattern to have sufficient grammar and syntax so as to itself constitute a ergodic information source Y. The output of the cognitive system, B, is deterministically backtranslated into the 'language' of Y, and we call that translation $\hat{Y}$. The rate distortion behavior relating Y and $\hat{Y}$, is, according to the RDT, determined by the mutual information $I(Y, \hat{Y})$. We take the index of coupling between the sensory input and the cognitive system to be the number of training cycles – an exposure measure – having an inverse K, and write

$$I(Y, \hat{Y}) = I[K] \tag{23}$$

$I[K]$ defines the splitting criterion between high and low probability pairs of training and response paths for a specified average distortion D, and is

analogous to the parametized information source uncertainty upon which we imposed renormalization symmetry to obtain phase transition.

We thus interpret the sudden changes in the measured average distortion $D \equiv \sum p(y)d(y,\hat{y})$ which determines 'mean square error' between training pattern and output pattern, e.g. the *ending* of a learning plateau, as representing onset of a phase transition in $I[K]$ at some critical K_C, consonant with our earlier developments.

Note that $I[K]$ constitutes an interaction between the cognitive system and the impinging sensory activity, so that its properties may be quite different from those of the cognitive condensation itself.

From this viewpoint learning plateaus are an inherently 'natural' phase transition behavior of pattern recognition systems. While one may perhaps, in the sense of [22], find more efficient gradient learning algorithms, our development suggests learning plateaus will be both ubiquitous and highly characteristic of a cognitive system. Indeed, it seems likely that proper analysis of learning plateaus will give deep insight into the structures underlying that system.

This is not a new thought: Mathematical learning models of varying complexity have been under constant development since the late 1940's [62], and learning plateau behavior has always been a focus of such studies.

The particular contribution of our perspective to this debate is that the distinct coevolutionary condensation of immune, CNS, and local sociocultural network cognition which distinguishes human biology must respond as a composite in a coherent, unitary and coupled manner to sensory input. Thus the 'learning curves' of the immune system, the CNS and the embedding sociocultural network are inevitably coupled and must reflect each other. Such reflection or interaction will, of necessity, be complicated.

Our analysis, however, has a particular implication. Learned cultural behavior – sociocultural cognition – is, from our viewpoint, a nested hierarchy of phase transition learning plateaus which carries within it the history of an individual's embedding socioculture. Through the cognitive condensation which distinguishes human biology, that punctuated history becomes part of individual cognitive and immune function. Simply removing 'constraints' which have deformed individual and collective past is unlikely to have the desired impact: one never, really, forgets how to ride a bicycle, and a social group, in the absence of affirmative redress, will not 'forget' the punctuated adaptations 'learned' from experiences of slavery or holocaust. Indeed, at the individual level, sufficiently traumatic events may become encoded within the

CNS and immune systems to express themselves as Post Traumatic Stress Disorder.

Noncognitive condensation in response to selection pressure

As discussed above, sociocultural networks serve multiple functions and are not only decision making cognitive structures, but are cultural repositories which embody the history of a community. Sociocultural networks, like human biology in the large, and the immune system in the small, have a duality in that they make decisions based on recognizing patterns of opportunity and threat by comparison with an internalized picture of the world, and they respond to selection pressure in the sense that cultural patterns which cannot adapt to external selection pressures simply do not survive. This is not learning in the traditional sense of neural networks. Thus the immune system has both 'innate' genetically programmed and 'learned' components, and human biology in the large is a convolution of genetic and cultural systems of information transmission.

We suggest that sociocultural networks – the instrumentalities of culture – likewise contain both cognitive and selective systems of information transmission which are closely intertwined to create a composite whole.

We now examine processes of 'punctuated evolution' inherent to evolutionary systems of information transmission.

We suppose a self-reproducing cultural system – more specifically a linked, and in the large sense coevolutionary, condensation of several such systems – is exposed to a structured pattern of selective environmental pressures to which it must adapt if it is to survive. From that adaptive selection – changes in genotype and phenotype analogs – we can infer, in a direct manner, something, but not everything, of the form of the structured system of selection pressures. That is, the culture contains markers of past 'selection events'.

We suppose the system of selection pressures to have sufficient internal structure – grammar and syntax – so as to itself constitute an ergodic information source Y whose probabilities are fixed on the timescale of analysis. The output of that system, B, is backtranslated into the 'language' of Y, and we call that translation $\hat{Y}$. The rate distortion behavior relating Y and $\hat{Y}$, is, according to the RDT, determined by the mutual information $I(Y, \hat{Y})$.

We take there to be a measure of the 'strength' of the selection pressure, P, which we use as an index of coupling with the culture of interest, having an inverse $K = 1/P$, and write

$$I(Y, \hat{Y}) = I[K].$$

(24)

P might be measured by the rate of attack by predatory colonizers, or the response to extreme environmental perturbation, and so on.

$I[K]$ thus defines the splitting criterion between high and low probability pairs of input and output paths for a specified average distortion D, and is analogous to the parametized information source uncertainty upon which we imposed renormalization symmetry to obtain phase transition. The result is robust in the absence of a distortion measure through the joint asymptotic equipartition theorem, as discussed above.

We thus interpret the sudden changes in the measured average distortion $D \equiv \sum p(y)d(y, \hat{y})$ which determines 'mean error' between pressure and response, i.e. the *ending* of a 'learning plateau', as representing onset of a phase transition in $I[K]$ at some critical K_C, consonant with our earlier developments. In the absence of a distortion measure, we may still expect phase transition in $I[K]$, according to the joint AEPT.

Note that $I[K]$ constitutes an interaction between the self-reproducing system of interest and the impinging ecosystem's selection pressure, so that its properties may be quite different from those of the individual or conjoined subcomponents [20, 21].

From this viewpoint highly punctuated 'non-cognitive condensations' are an inherently 'natural' phase transition behavior of evolutionary systems, even in the absence of a distortion measure. Again, while there may exist, in the sense of [59], more efficient convergence algorithms, our development suggests plateaus will be both ubiquitous and highly characteristic of evolutionary process and path. Indeed, it seems likely that proper analysis of non-cognitive evolutionary 'learning' plateaus – to the extent they can be observed or reconstructed – will give deep insight into the mechanisms underlying that system.

Convolution between selection pressure and sociocultural cognition

Selection pressure acting on sociocultural networks can be expected to affect their cognitive function, their ability to recognize and respond to relatively immediate patterns of threat and opportunity. In fact, those patterns themselves may in no small part represent factors of that selection pressure, conditionally dependent on it. We assume, then, the linkage of *three* information sources, two of which are conditionally dependent on and may indeed be dominated by, a highly structured embedding system of externally imposed selection pressure which we call Y_3. Y_2 we will characterize as the pattern recognition-and-response language of the sociocultural network itself.

In IR Cohen's [17-19] sense, this involves comparison of sensory information with an internalized picture of the world, and choice of a response from a repertory of possibilities. Y_1 we take to be a more rapidly changing, but nonetheless structured, pattern of immediate threat-and-opportunity which demands appropriate response and resource allocation – the 'training pattern'. We reiterate that Y_1 is likely to be conditionally dependent on the embedding selection pressure, Y_3, as is the hierarchically layered history expressed by Y_2.

According to the triplet version of the SMT which we discussed at the end of the theoretical section above, then, for large n, triplets of paths in Y_1, Y_2 and Y_3 may be divided into two sets, a smaller 'meaningful' one of high probability – representing those paths consistent with the 'grammar' and 'syntax' of the interaction between the selection pressure, the cognitive sociocultural process, and the pattern of immediate 'sensory challenge' it faces – and a very large set of vanishingly small probability. The splitting criterion is the conditional mutual information:

$$I(Y_1, Y_2|Y_3).$$

We parametize this splitting criterion by a variate K representing the inverse of the strength of the coupling between the system of selection pressure and the linked complex of the sociocultural cognitive process and the structured system of day-to-day problems it must address. $I[K]$ will, according to the 'phase transition' developments above, be highly punctuated by 'mixed' plateau behavior representing the synergistic and inextricably intertwined action of both externally imposed selection pressure and internal sociocultural cognition.